PATENT
2786-203P

IN THE U.S. PATENT AND TRADEMARK OFFICE

Applicant:	Zeev Smilansky	Conf.:	5637
Appl. No.:	10/056,483	Group:	2621
Filed:	January 28, 2002	Examiner:	UNKNOWN
For:	METHOD FOR COMPARING SIGNAL ARRAYS IN DIGITAL IMAGES		

#4
6-25-02

L E T T E R

Assistant Commissioner for Patents
Washington, DC 20231

June 4, 2002

Sir:

Under the provisions of 35 U.S.C. § 119 and 37 C.F.R. § 1.55(a), the applicant(s) hereby claim(s) the right of priority based on the following application(s):

Country	Application No.	Filed
ISRAEL	141151	January 29, 2001

A certified copy of the above-noted application(s) is(are) attached hereto.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fee required under 37 C.F.R. §§ 1.16 or 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By _____
Charles Gorenstein, #29,271

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

CG/RWD/sjl
2786-203P

Attachment

2786-203P
10/058,483
1-24-02
Zeev Smilansky
Birch, Stewart,
Kolasch & Birch
703-205-8000

מדינת ישראל
STATE OF ISRAEL

istry of Justice
atent Office

משרד המשפטים
לשכת הפטנטים

to certify that annexed
a true copy of the
as originally
with the patent
of which
specified on the
annex.

זאת לתעודה כי רצופים
בזה העתקים נכונים של
המסמכים שהופקדו
לכתחילה עם הבקשה
לפטנט לפי הפרטים
הרשומים בעמוד הראשון
של הנספח.

This 08-05-2002 היום

רשם הפטנטים
Commissioner of Patents

CERTIFIED COPY OF PRIORITY DOCUMENT

נתאשר
Certified

חוק הפטנטים, תשכ"ז – 1967
PATENTS LAW, 5727-1967

בקשה לפטנט
Application For Patent

לשימוש הלשכה
For Office Use

מספר: Number	141151
תאריך: Date	23-01-2001
הוקדם/נדחה: Ante/Post-dated	

אני, (שם המבקש, מענו ולגבי גוף מאוגדת מקום התאגדותו)
I, (Name and address of applicant, and in case of body corporate-place of incorporation)

קומפיוגן בע"מ, חברה ישראלית מרחוב פנחס רוזן 72, תל אביב 69512, ישראל
Compugen Ltd., Israeli Company of 72 Pinchas Rosen Street, Tel Aviv 69512, ISRAEL

בעל אמצאה מכח: Owner, by virtue of — הדין: Right of Law

ששמה הוא: of an invention the title of which is

(בעברית) (Hebrew)

שיטה להשוואת מערכי אותות בתמונות ספרתיות

(באנגלית) (English)

Method for comparing signal arrays in digital images

מבקש בזאת כי ינתן לי עליה פטנט
Hereby apply for a patent to be granted to me in respect thereof.

* בקשת חלוקה Application of Division	* בקשת פטנט מוסף Appl. for Patent of Addition	דרישת דין קדימה* Priority Claim		
מבקשת פטנט from application	* לבקשה/לפטנט to Patent/Appl.	מספר/סימן Number/Mark	תאריך Date	מדינת האיגוד Convention Country
מס' No.	מס' No.			
מיום Dated	מיום Dated			
* יפוי כח: כללי P.o.A.: General				
הוגש בעניין filed in case P128131				

המען למסירת מסמכים בישראל
Address for Service in Israel

REINHOLD COHN AND PARTNERS
Patent Attorneys
P.O.B. 4060, Tel-Aviv
C. 125624

חתימת המבקש
Signature of Applicant

For the Applicants,
REINHOLD COHN AND PARTNERS
By: — Ben Zeevi ✓

This היום 28 בחודש January שנת 2001
Year of

לשימוש הלשכה
For Office Use

טופס זה כשהוא מוטבע בחותם לשכת הפטנטים ומושלם במספר ובתאריך ההגשה, הנו אישור להגשת הבקשה שפרטיה רשומים לעיל.
This form, impressed with the Seal of the Patent Office and indicating the number and date of filing, certifies the filing of the application the particulars of which are set out above.

* מחק את המיותר
Delete whatever is inapplicable

שיטה להשוואת מערכי אותות בתמונות ספרתיות

Method for comparing signal arrays in digital images

קומפיוגן בע"מ

Compugen Ltd.

C. 125624

METHOD FOR COMPARING SIGNAL ARRAYS IN DIGITAL IMAGES.

FIELD OF THE INVENTION

The invention relates to methods of comparing the intensity of two signal arrays in digital images, for example digital images of a spot in a one- or two-dimensional electrophoresis pattern or a DNA chip.

BACKGROUND OF THE INVENTION

A digital image may be considered to be an array of signals, where each pixel in the image produces a visible signal of a particular intensity. It is often of interest to compare two such signal arrays. For example, two protein mixtures can be separated by one of various separation techniques to produce two one- or two-dimensional separation patterns. A digital image of a spot in each pattern, corresponding to the same protein could be compared in order to compare the amount of the protein present in each mixture. As another example, a DNA chip having attached to it various oligonucleotide targets is incubated in the presence of probe oligonucleotides from two sources. The two probe species are differently labeled, so that each probe species produces a visible signal that is distinguishable from that of the other species. For example, one probe species may be labeled with a fluorescent dye that produces a red signal while the other probe species is labeled with a fluorescent dye that produces a green signal. A digital image of the red signal could then be compared with a digital image of the green signal in order to compare the amount of oligonucleotides binding to the chip in the two sources.

One well-known method for comparing the signal arrays in two digital images involves calculating the total intensity in each image and then calculating

Another method is to determine the maximum
ratio of the two maximal intensities.

method for comparing two visual signal
ple, a digital image of a stained spot in a
tern such as produced by electrophoresis.
mage of a region of a DNA chip that has
iat produce a visible signal. The two arrays
eparated from one another or superimposed

ivention, the two signal arrays to be compared
r. The two arrays may be, for example, a single
NA chip that was simultaneously incubated with
different sources, where the probes from each
r producing a distinct visible signal. For example,
ay be labeled with a fluorescent label producing a
m the other source labeled with a label producing a
red and green signal arrays in the digital image are
other, and are to be compared by the method of the

/s are superimposed upon one another, each pixel x_i in
escribed by an ordered pair of numbers $(I_1(x_i), I_2(x_i))$
ity of the signal of the pixel x_i in the first array, and $I_2(x_i)$
gnal of the pixel x_i in the second array. A linear regression
he points $(I_1(x_i), I_2(x_i))$. Within the context of the present
inear regression" is used to include any method in which a
a set of points, for example, a least squares fit of the points
n in the art. This also includes methods involving a filtering
ts are deleted from the set of points prior to determining the

invention, the two arrays are compared by
ed by the linear regression analysis.
invention, two signal arrays are compared
another. The two patterns may be, for
erent one- or two- dimensional separation
oresis. The two arrays are first put into
he two patterns is described by means of
the first pattern to a pixel $\mathbf{T}(x_i)$ in the
istration transformations are disclosed,
3562 Two arrays in register with each
red in accordance with the invention
, an ordered pair of numbers $(I(x_i)$,
ty of the signal of a pixel x_i in the
el $\mathbf{T}(x_i)$ in the second pattern that
n analysis is applied to the points
the two arrays are compared by
luced by the linear regression

nination of differential gene
nal arrays to be compared
gene. Typically, but not
he gene expression under
for the determination of
of the signal arrays to be
esent in a sample.

may be carried out in
way of non-limiting
which:

linear fit. In accordance with the invention, the two arrays are compared by means of the slope of the line produced by the linear regression analysis.

In another embodiment of the invention, two signal arrays are compared that are not superimposed upon one another. The two patterns may be, for example, digital images of spots in different one- or two- dimensional separation patterns such as produced by electrophoresis. The two arrays are first put into register with each other. Registration of the two patterns is described by means of a transformation **T** that maps a pixel x_i in the first pattern to a pixel $\mathbf{T}(x_i)$ in the second pattern. Methods for obtaining registration transformations are disclosed, for example, in Israel Patent Application 133562 Two arrays in register with each other under the transformation **T** are compared in accordance with the invention as follows. For each pixel x_i in the first array, an ordered pair of numbers $(I(x_i), I(\mathbf{T}(x_i))$ is generated where $I(x_i)$ is the intensity of the signal of a pixel x_i in the first array and $I(\mathbf{T}(x_i))$ is the intensity of the pixel $\mathbf{T}(x_i)$ in the second pattern that is in register with the pixel x_i. A linear regression analysis is applied to the points $(I(x_i), I(\mathbf{T}(x_i))$. In accordance with the invention, the two arrays are compared by means of the slope of the regression line produced by the linear regression analysis.

The invention may be used for the determination of differential gene expression. In this application, each of the signal arrays to be compared represents the level of expression of a particular gene. Typically, but not necessarily, the two arrays represent the level of the gene expression under different conditions. The invention may also be used for the determination of differential protein expression. In this application, each of the signal arrays to be compared represents the amount of a particular protein present in a sample.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to understand the invention and to see how it may be carried out in practice, a preferred embodiment will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Fig. 1 is a plot of the ordered pairs $(I_1(x), I_2(x_i))$ where $I_1(x_i)$ is the intensity of a signal produced by a first DNA probe species in the pixel x_i, $I_2(x_i)$ is the intensity of a signal produced by a second DNA probe species in the pixel x_i, the DNA probes being bound to DNA targets on a DNA chip;

Fig. 2 shows two two-dimensional separation patterns;

Fig. 3 shows a enlargement of first and second spots from the first and second separation patterns, respectively, of Fig. 2, and

Fig. 4 shows a plot of the points $(I(x_i),\mathbf{T}(I(x_i)))$, where $I(x_i)$ is in the intensity of a pixel x_i in the first spot of Fig. 3 and $I(\mathbf{T}(x_i))$ is the intensity of a pixel $\mathbf{T}(x_i)$ in the second spot that is in register with the first spot under a transformation **T**.

EXAMPLES

Example 1 Two superimposed spots

A DNA chip having DNA targets bound on it was incubated in the presence of a sample containing first and second DNA probe species, where each probe species was labeled with a label producing a distinct visible signal. Each of the first and second probe species bound to a particular target on the chip thus produces a distinct signal array in a region of the chip where the target is located. For a pixel x_i, the intensity of the two signal arrays is represented by an ordered pair of numbers $(I_1(x_i), I_2(x_i))$ where $I_1(x_i)$ is the intensity of the signal produced by the first probe species in the pixel x_i and $I_2(x_i)$ is the intensity of the signal produced by the second probe species in the pixel x_i. Fig. 1 shows a plot of the ordered pairs $(I_1(x_i), I_2(x_i))$. A linear regression analysis was applied to the points $(I_1(x_i), I_2(x_i))$ that produced the best linear fit **200** to the points. The slope of the line **200** was found to be 1.48, indicating that a probes of the second species binding to a particular target on the chip were present in the sample at an abundance of about 1.48 times that of probes of the first species binding to the same target. The two spots are compared by means of the slope of the line **200.**

Example 2 Separated arrays

Two samples containing proteins are separated to produce a pair of two-dimensional separation patterns. Fig. 2 shows a representation of two

two-dimensional separations patterns **305** and **310.** A spot **315** in the first pattern **305** is to be compared with a spot **320** in the second pattern **310.** Fig. 3 shows enlargements of the spots **315** and **320,** divided into pixels. The pixels in each spot form a signal array. Each pixel in the spot **315,** for example, the pixel **325** has an associated intensity $I(x_i)$. Similarly, each pixel y_i in the spot **320,** for example the spot **330,** has an associated intensity $I(y_i)$. A mapping **T** is found that maps each of a plurality of pixels in the spot **315** to a different pixel in the spot **320.** For example, the pixel **325** may be mapped into the pixel **330.**

If the two spots **315** and **320** consist of the same number of pixels, then the mapping **T** may be obtained by first putting the entire patterns **305** and **310** into register with each other. The patterns **305** and **310** are put in register with one another by means of a transformation **T** that maps each pixel x_i in the pattern **305,** for example the pixel **330** to a pixel $\mathbf{T}(x_i)$ in the pattern **310.** A transformation that puts the two patterns into register with each other may be found, for example, as disclosed in Israel Patent Application No. 133562. The restriction of the transformation **T** to the spot **315** maps pixels in the spot **315** to pixels in the spot **320.**

Another method that may be used to put the spots **315** and **320** into register with each other when the two spots consist of about the same number of pixels is to arrange the pixels in each spot in order of decreasing intensity. The mapping **T** is then defined that maps the nth pixel in the arrangement of the pixels of the spot **315** with the nth spot in the arrangement of the pixels of the spot **320.**

When the two spots **315** and **320** consist of about the same number of pixels, and the mapping **T** has been defined, pairs of numbers are $(I(x_i), I(\mathbf{T}(x_i)))$ formed where $I(x_i)$ is in the intensity of a pixel x_i in the pattern **105** and $I(\mathbf{T}(x_i))$ is the intensity of the pixel $\mathbf{T}(x_i)$ in the pattern **115** that is in register with x_i under the transformation **T**. Fig. 4 shows a plot of the points $(I(x_i),\mathbf{T}(I(x_i)))$. A linear regression analysis is applied to the points that produces the best linear fit **400** to the points. The slope of the linear fit **400** is found to be 4.8 indicating that the spot

320 contains about 4.8 as much protein as is present in the spot **315.** The two spots are compared by means of the slope of the line **400.**

If, say, the spot **315** consists of substantially more pixels than the spot **320,** the following method may be used to put a plurality of the pixels of the spot **315** into register with pixels in the spot **320.** The pixels in each spot are arranged in order of decreasing intensity. A predetermined fraction r_1 of the pixels in the spot **315** are then deleted from the arrangement of the pixels of that spot, to produce a provisional arrangement of the pixels of that spot. A predetermined fraction r_2 of the pixels in the spot **320** are then deleted from the arrangement of the pixels of that spot, to produce a provisional arrangement of the pixels of that spot. r_1 and r_2 are selected so that the two provisional arrangements consist of about the same number of pixels. Preferably, the pixels deleted to form the provisional arrangements are substantially uniformly distributed in each of the initial arrangements. Thus, about every $1/r_1$-th pixel is removed from the initial sequence of pixels from the spot **315** and about every $1/r_2$-th pixel is removed from the initial sequence of pixels from the spot **320.** A transformation **T'** is then defined that maps the nth pixel in the provisional arrangement of the pixels of the spot **315** with the nth spot in the provisional arrangement of the spot **320.**

Pairs of numbers are $(I(x), I(\mathbf{T}'(x_i)))$ formed where $I(x_i)$ is in the intensity of a pixel x_i in the pattern **105** and $I(\mathbf{T}'(x_i))$ is the intensity of the pixel $\mathbf{T}'(x_i)$ in the pattern **115** that is in register with x under the transformation **T'**. Fig. 5 shows a plot of the points $(I(x_i).\mathbf{T}'(I(x_i)))$. A linear regression analysis is applied to the points that produces the best linear fit **500** to the points. The slope of the linear fit **500** is multiplied by r_2/r_1 to compensate for the deletion of points from the two spot arrangements.

It will also be understood that the system according to the invention may be a suitably programmed computer. Likewise, the invention contemplates a computer program being readable by a computer for executing the method of the invention. The invention further contemplates a machine-readable memory tangibly

embodying a program of instructions executable by the machine for executing the method of the invention.

CLAIMS:

1. A method for comparing first and second signal arrays, the arrays being comprised of pixels, each pixel in an array having an intensity, the method comprising steps of:

(a) associating to each of a plurality of pixels x_i in the first array a pixel $\mathbf{T}(x_i)$ in the second array, and

(b) applying a linear regression analysis to the ordered pairs of numbers $(x_i, T(x_i))$ so as to produce a slope.

2. The method according to Claim 1 wherein the first and second signal arrays are superimposed and $T(x_i)=x_i$.

3. The method according to Claim 2 wherein the first and second signal arrays are obtained by incubating a DNA chip in the presence of first and second probe species, the first probe species producing a signal that is distinguishable from a signal produced by the second probe species.

4. The method according to Claim 2 wherein the first and second signal arrays are obtained by staining a spot in separation pattern with first and second labels, the first label producing a signal that is distinguishable from a signal produced by the second label.

5. The method according to Claim 1 wherein the first and second arrays are not superimposed.

6. The method according to Claim 5 wherein the first and second signal arrays are spots in a first and second separation pattern, respectively.

7. The method according to Claim 6 wherein the first and second separation patterns are in register, and for each pixel x_i in the first spot, $\mathbf{T}(x_i)$ is the spot in the second separation pattern in register with x_i.

8. The method according to any one of the previous claims for use in determining differential gene expression or differential protein expression.

9. A method for determining differential gene expression of a gene comprising steps of:

(a) obtaining digitized images of first and second signal arrays representing first and second expression levels of the gene, respectively, each pixel in an image having an intensity;

(b) associating to each of a plurality of pixels x_i in the first image a pixel $\mathbf{T}(x_i)$ in the second image, and

(c) applying a linear regression analysis to the ordered pairs of numbers $(x_i, \mathbf{T}(x_i))$ so as to produce a slope.

10. The method according to Claim 9 wherein the first and second signal arrays are superimposed and $T(x_i)=x_i$.

11. The method according to Claim 10 wherein the first and second signal arrays are obtained by incubating a DNA chip in the presence of first and second probe species, the first probe species producing a signal that is distinguishable from a signal produced by the second probe species.

12. The method according to Claim 10 wherein the first and second signal arrays are obtained by staining a spot in separation pattern with first and second labels, the first label producing a signal that is distinguishable from a signal produced by the second label.

13. A method for determining differential protein expression comprising steps of:

(a) obtaining digitized images of first and second signal arrays representing first and second expression levels of the protein, respectively, each pixel in an image having an intensity;

(b) associating to each of a plurality of pixels x_i in the first image a pixel $\mathbf{T}(x_i)$ in the second image, and

(c) applying a linear regression analysis to the ordered pairs of numbers $(x_i, \mathbf{T}(x_i))$ so as to produce a slope.

14. The method according to Claim 13 wherein the first and second arrays are not superimposed.

15. The method according to Claim 14 wherein the first and second signal arrays are spots in a first and second separation pattern, respectively.

16. The method according to Claim 15 wherein the first and second separation patterns are in register, and for each pixel x_i in the first spot, **T**(xi) is the spot in the second separation pattern in register with x_i.

17. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for comparing digitized images of first and second signal arrays, the images being comprised of pixels, each pixel in an image having an intensity, the method comprising steps of:

(a) associating to each of a plurality of pixels x_i in the first image a pixel **T**(x_i) in the second image, and

(b) applying a linear regression analysis to the ordered pairs of numbers (x_i, **T**(x_i)) so as to produce a slope.

18. A computer program product comprising a computer useable medium having computer readable program code embodied therein for comparing digitized images of first and second signal arrays, the images being comprised of pixels, each pixel in an image having an intensity, the computer program product comprising:

computer readable program code for causing the computer to associate to each of a plurality of pixels x_i in the first image a pixel **T**(x_i) in the second image, and

computer readable program code for causing the computer to apply a linear regression analysis to the ordered pairs of numbers (x_i, T(x_i)) so as to produce a slope.

19. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for determining differential gene expression of a gene comprising steps of:

(a) obtaining digitized images of first and second signal arrays representing first and second expression levels of the gene, respectively, each pixel in an image having an intensity;

(b) associating to each of a plurality of pixels x_i in the first image a pixel **T**(x_i) in the second image, and

(c) applying a linear regression analysis to the ordered pairs of numbers (x_i, $\mathbf{T}(x_i)$) so as to produce a slope.

20. A computer program product comprising a computer useable medium having computer readable program code embodied therein for determining differential gene expression of a gene the computer program product comprising:

computer readable program code for causing the computer to obtain digitized images of first and second signal arrays representing first and second expression levels of the gene, respectively, each pixel in an image having an intensity;

computer readable program code for causing the computer to associate to each of a plurality of pixels x_i in the first image a pixel $\mathbf{T}(x_i)$ in the second image, and

computer readable program code for causing the computer to apply a linear regression analysis to the ordered pairs of numbers (x_i, $\mathbf{T}(x_i)$) so as to produce a slope.

21. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for determining differential protein expression comprising steps of:

(a) obtaining digitized images of first and second signal arrays representing first and second expression levels of the protein, respectively, each pixel in an image having an intensity;

(b) associating to each of a plurality of pixels x_i in the first image a pixel $\mathbf{T}(x_i)$ in the second image, and

(c) applying a linear regression analysis to the ordered pairs of numbers (x_i, $\mathbf{T}(x_i)$) so as to produce a slope.

22. A computer program product comprising a computer useable medium having computer readable program code embodied therein for determining differential protein expression the computer program product comprising:

computer readable program code for causing the computer to obtain digitized images of first and second signal arrays representing first and second

expression levels of the protein, respectively, each pixel in an image having an intensity;

computer readable program code for causing the computer to associate to each of a plurality of pixels x_i in the first image a pixel $\mathbf{T}(x_i)$ in the second image, and

computer readable program code for causing the computer to apply a linear regression analysis to the ordered pairs of numbers $(x_i, \mathbf{T}(x_i))$ so as to produce a slope.

For the Applicants,
REINHOLD COHN AND PARTNERS
By:

Ben Spungin

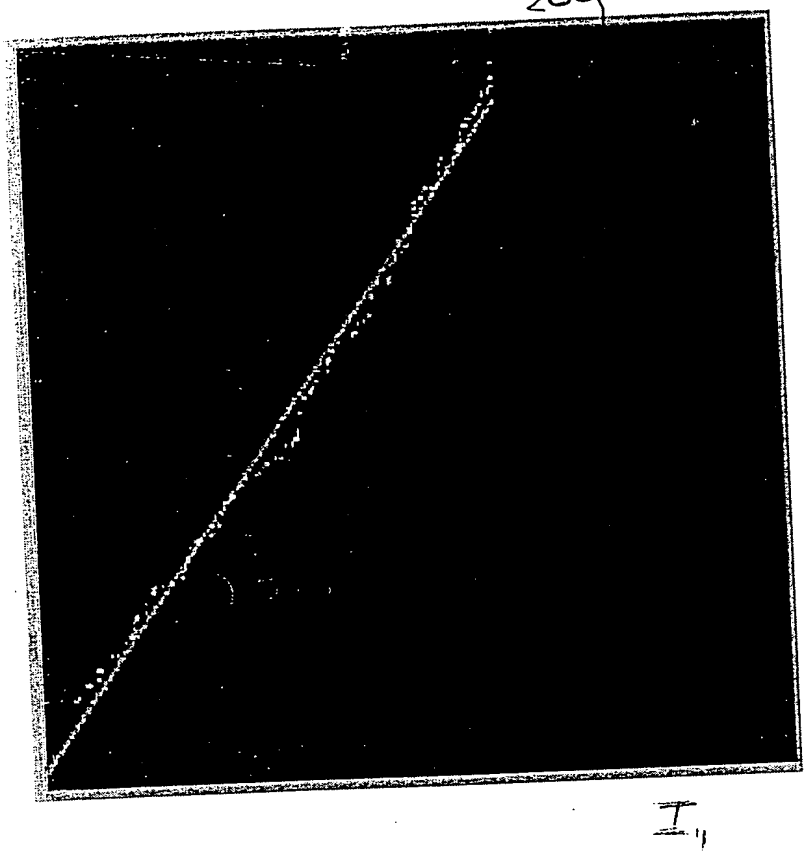

FIG. 1

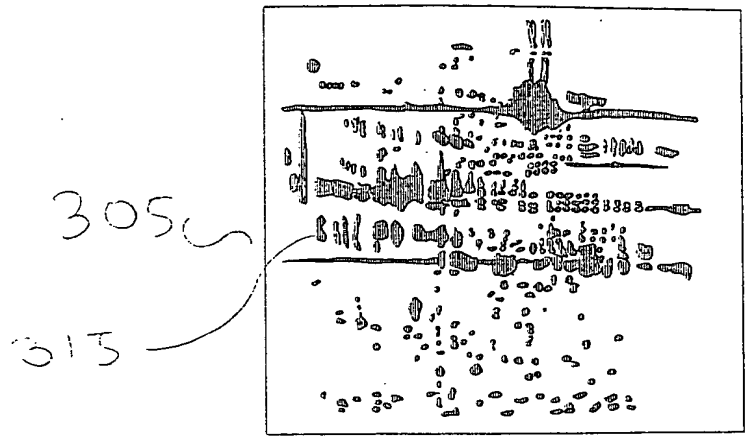

FIG. 1A

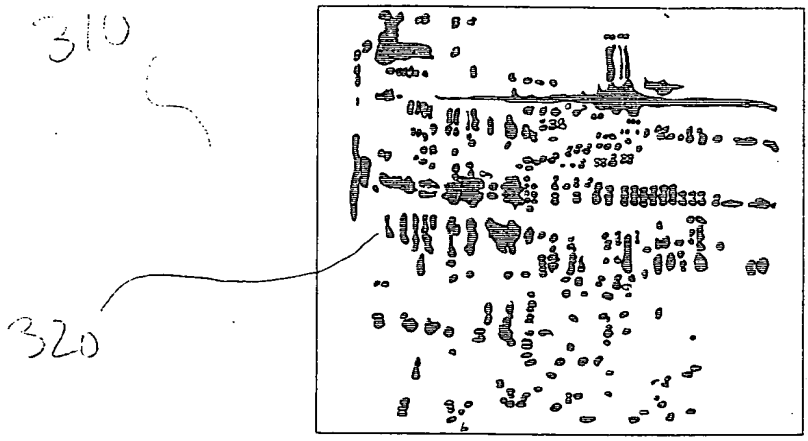

FIG. 1B

FIG. 2

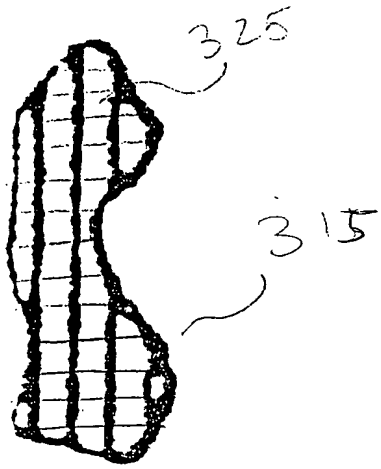

FIG. 3

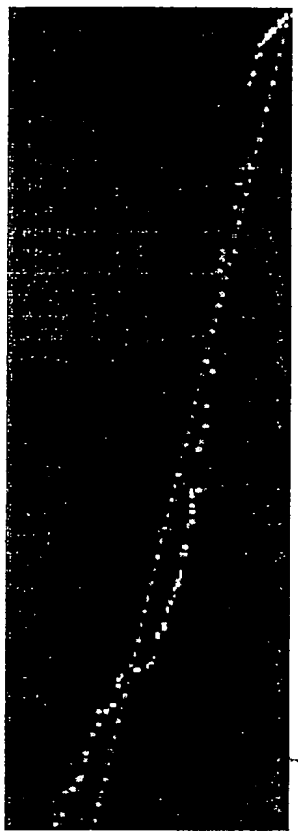

FIG. 4